Statistics 403 Problem Set 7

Due in lab on Friday, October 30th

1. Suppose we collect paired data X_i, Y_i, in which the two values represent measures of a certain disease symptom in a given individual before and after the patient is treated. Suppose that the unstandardized effect size is $\delta = EY - EX = 0.1$, the variances are $\mathrm{var}(X) = \mathrm{var}(Y) = 1$, and the correlation between X_i and Y_i is 0.4.

   For both parts, you can use $P(T > 2)$ alone as the power (i.e. ignore $P(T < -2)$, and assume that $\delta > 0$).

   (a) What sample size is required to get 80% power if the data are analyzed using a paired test?

   Solution: Denote the test statistic by $T = \sqrt{n}\bar{D}/\hat{\sigma}_D$, and note that the standard deviation of the data is $\sigma_D = \sqrt{1 + 1 - 2 \cdot 0.4} \approx 1.1$. Using $P(T > 2)$ as an approximation to the power, we get

$$
\begin{aligned}
P(T > 2) &= P(\sqrt{n}\bar{D}/\hat{\sigma}_D > 2) \\
&\approx P(\sqrt{n}\bar{D}/1.1 > 2) \\
&= P(\sqrt{n}(\bar{D} - 0.1)/1.1 > 2 - \sqrt{n} \cdot 0.1/1.1) \\
&= P(Z > 2 - \sqrt{n} \cdot 0.09).
\end{aligned}
$$

   To get 80% power we need $2 - \sqrt{n} \cdot 0.09$ to equal the $1 - 0.8 = 0.2$ quantile of the standard normal distribution, which is -0.84. Solving $2 - \sqrt{n} \cdot 0.09 = -0.84$ yields $n = 996$.

   (b) What sample size is required to get 80% power if the data are analyzed using an unpaired test? Hint: you will need to compensate for the fact that the wrong standard deviation is being used in the test statistic.

   Solution: If we treat the data as independent samples, we will get the following power as a function of the sample size n:

$$
\begin{aligned}
P(T > 2) &= P(\bar{D}/\sqrt{\hat{\sigma}_x^2/n + \hat{\sigma}_y^2/n} > 2) \\
&\approx P(\sqrt{n}\bar{D}/\sqrt{2} > 2) \\
&= P(\sqrt{n}(\bar{D} - 0.1)/\sqrt{2} > 2 - \sqrt{n} \cdot 0.1/\sqrt{2}) \\
&= P(\sqrt{n}(\bar{D} - 0.1)/1.1 > 2\sqrt{2}/1.1 - \sqrt{n}0.1/1.1) \\
&= P(Z > 2.57 - 0.09\sqrt{n}).
\end{aligned}
$$

   To get the sample size, as above we set $2.57 - 0.09\sqrt{n} = -0.84$ and solve for n to get $n = 1436$.

Note that a much greater sample size is needed because we are using the data inefficently. We are carrying out the test using $\mathrm{SD}(D_i) = 1.4$, when in fact $\mathrm{SD}(D_i)$ is only 1.1, due to the positive correlation between X_i and Y_i.